

# Effect of Training Rate on Recognition of Spectrally Shifted Speech

Geraldine Nogaki, Qian-Jie Fu, and John J. Galvin III

**Objective:** Previous studies have shown that the protocol used for auditory training may significantly affect the outcome of training. However, it is unclear how often training should be performed to maximize its benefit. The present study investigated how the frequency of training contributed to normal-hearing listeners' adaptation to spectrally shifted speech.

**Methods:** Eighteen normal-hearing listeners were trained with spectrally shifted and compressed speech via an 8-channel acoustic simulation of cochlear implant speech processing. Five short training sessions (1 hr per session) were completed by each subject; subjects were trained at one of three training rates: five sessions per week, three sessions per week, or one session per week. Subjects were trained to identify medial vowels presented in a cVc format; depending on the level of difficulty, the number of response choices was increased and/or the acoustic differences between vowels were reduced. Vowel and consonant recognition was measured before and after training as well as at regular intervals during the training period. Sentence recognition was measured before and after training only.

**Results:** Results showed that pretraining vowel recognition scores were poor (14.0% correct, on average) for all subjects, due to the severe spectral shift. After five sessions of targeted vowel contrast training, there was a significant improvement of shifted vowel recognition for most subjects. The mean improvement was comparable (~15 percentage points) across the three training rate conditions, despite significant intersubject variability in pre- and pretraining baseline performance. There was no significant difference in training outcomes among the three training rates. Spectrally shifted consonant and sentence recognition also improved by ~20 percentage points after training, even though consonants and sentences were not explicitly trained. Similar to vowel recognition, there was no significant difference in training outcomes among the three training rates for shifted consonant and sentence recognition.

**Conclusions:** The results demonstrated that the training rate had little effect on normal-hearing listeners' adaptation to spectrally shifted speech, at least for the training periods (ranging from 1 to 5 wk) used in the present study. The outcome of auditory training may depend more strongly on the amount of training (i.e., total number of training sessions) rather

than the frequency of training (i.e., daily or once per week). Although more frequent training may accelerate listeners' adaptation to spectrally shifted speech, there may be significant benefits from training as little as one session per week. The results of the present study suggest that appropriate training schedules can be developed to optimize the effectiveness, efficiency, and effort associated with hearing-impaired patients' auditory rehabilitation.

(*Ear & Hearing* 2007;28;132-140)

With advances in cochlear implant technology, the overall speech recognition of patients with cochlear implants has steadily improved. With the most advanced implant device and speech processor, many patients with cochlear implants receive great benefit and are capable of conversing with friends and family over the telephone. However, considerable variability remains in individual patient outcomes. Some patients receive little benefit from the latest cochlear implant technology, even after many yr of daily use of the device. Although considerable efforts have been made to develop and optimize speech processing strategies for poorly performing patients, auditory training is also an important approach toward improving speech recognition performance in patients with cochlear implants (Fu et al., 2005a).

Previous studies have shown mixed results with auditory training for poorly performing patients with cochlear implants. Busby et al. (1991) measured the effect of auditory training on the speech recognition performance of three prelingually deafened cochlear implant users (two adults and one adolescent). Auditory training consisted of 10, 1-hr sessions (one to two sessions per week). After training, there were only minimal changes in these subjects' speech performance; the subject who improved the most was implanted at an earlier age than the other two subjects and therefore had a shorter period of deafness. Dawson & Clark (1997) later investigated the effects of auditory training on the vowel recognition performance of five congenitally deafened patients (three children, one adolescent, and one young adult). Training was specifically focused on improving vowel perception. Training was provided once per week for 10 wk, for a total of 10 training sessions; each training session lasted approximately 50 minutes. Results showed that after training, two children showed significant gains on a number of tests; however, there

Department of Auditory Implants and Perception, House Ear Institute, Los Angeles, California.

were only minimal improvements for the remaining three subjects with cochlear implants. Recently, Fu et al. (2005a) investigated the effects of auditory training on the speech recognition performance in seven prelingually deafened and three postlingually deafened adult patients with cochlear implants who have limited speech recognition abilities. Subjects were trained with their home computers using speech stimuli and custom training software; subjects trained 1 hr per day, 5 d per week, for a period of 1 mo or longer. Using monosyllabic words, subjects were trained to identify medial vowels. Auditory and visual feedback was provided, allowing subjects to repeatedly compare their (incorrect) choice to the correct response. Results showed a significant improvement in all subjects' speech perception performance after this moderate but regular training.

The type of training protocol (i.e., phonetic contrast training) used in Fu et al. (2005a) may have contributed to the better training outcomes than those observed with previous studies. Fu et al. (2005b) investigated the effect of different training protocols on 16 normal-hearing (NH) subjects' ability to learn spectrally shifted speech; all training and testing was conducted by using spectrally shifted speech. Short daily training sessions were conducted over 5 consecutive days, using three different training protocols and one test-only protocol. Subjects in the test-only protocol received no preview, no feedback, and no training. Subjects in the "preview" protocol were asked to preview the 12 hVd tokens used in the vowel recognition test before each test. Subjects in the "vowel contrast training" protocol were trained to identify medial vowels using monosyllabic words in a cVc context. Subjects in the "sentence training" protocol were trained by using modified connected discourse tracking (DeFilippo & Scott, 1978), similar to methods used in previous training studies (Fu & Galvin, 2003; Rosen et al., 1999). Results showed that recognition of spectrally shifted vowels was significantly improved by training with the preview and vowel protocols; no significant improvement in vowel recognition was observed with the test-only or sentence training protocols. These results suggest that training protocols may significantly contribute to auditory training outcomes.

Another factor that may contribute to differences in auditory training outcomes is the amount of training. For example, in the previous Busby (1991) and Dawson & Clark (1997) studies, subjects completed 10 training sessions; in the Busby study, subjects were trained 1 to 2 times per week, whereas in the Dawson & Clark study, subjects were trained once per week. In Fu et al. (2005a), subjects trained for 1 hr per day, 5 d per week for a period of 1 mo or longer, resulting in a minimum of 20 training sessions. Re-

cently, Wright and her colleagues explored patterns of learning and generalization on a variety of basic auditory tasks provided by different amounts of training (Fitzgerald & Wright, 2005; Ortiz et al., Reference Note 1). They found that less than 1 hr of training appeared to yield less learning but more generalization than did multihour training. Rosen et al. (1999) used connected discourse tracking (DeFilippo & Scott, 1978) to train listeners' recognition of four-channel, spectrally shifted speech. They found that performance improved significantly for vowel, consonant, and sentence recognition after just nine 20-minute sessions (3 hr) of connected discourse tracking with the shifted simulation. In addition, differences between previous studies in terms of the total number of training sessions/hours and the frequency of training may have contributed to differences in training outcomes.

As both a practical and theoretical consideration, it is important to understand the effects of the frequency of training on auditory training outcomes to design appropriate training schemes that patients with cochlear implants may integrate into their daily lives. For some busy patients, this may mean committing to the fewest number of training sessions per week possible while still getting some benefit. The present study explored the effects of training rate on auditory training outcomes. NH subjects were trained and tested with spectrally shifted speech while listening to an acoustic cochlear implant simulation similar to that used by Fu et al. (2005b). Subjects were trained using the vowel contrast protocol used in Fu et al. (2005a, 2005b). The total number of training sessions was fixed (5-hr-long sessions). Depending on the subject group, the frequency of training was one, three, or five sessions per week—training rates that are typical of those used in previous studies.

## MATERIALS AND METHODS

### A. Subjects

Eighteen NH adults (10 women and 8 men), ages 21 to 39 yr, participated in the study. All subjects had pure-tone thresholds better than 20 dB HL at octave frequencies ranging from 125 to 8000 Hz. All subjects were native speakers of American English. All subjects were paid for their participation.

### B. Signal Processing

Normal-hearing subjects were trained and tested while listening to 8-channel acoustic simulations of cochlear implant speech processing implemented with the Continuously Interleaved Sampling (CIS) strategy (Wilson et al., 1991). The sine-wave vocoders used in the cochlear implant simulations were

**TABLE 1. Corner frequencies of analysis/carrier filters and the frequencies of sine wave carriers used in the acoustic cochlear implant simulation**

Channel no. (apex to base)	Analysis band unshifted carrier band corner frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Center frequency of carrier filter (Hz)	Shifted carrier band corner frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Center frequency of carrier filter (Hz)
1	200–359	5.3–8.1	268	999–1,363	14–16	1,167
2	359–591	8.1–10.8	461	1,363–1,843	16–18	1,585
3	591–930	10.8–13.6	741	1,843–2,476	18–20	2,136
4	930–1,426	13.6–16.3	1,152	2,476–3,310	20–22	2,863
5	1,426–2,149	16.3–19.0	1,751	3,310–4,410	22–24	3,821
6	2,149–3,205	19.0–21.8	2,624	4,410–5,860	24–26	5,084
7	3,205–4,748	21.8–24.5	3,901	5,860–7,771	26–28	6,748
8	4,748–7,000	24.5–27.3	5,765	7,771–10,290	28–30	8,942

Corner frequencies of analysis filters and sinusoidal frequencies of sine wave carriers used in the sine wave cochlear implant simulation.

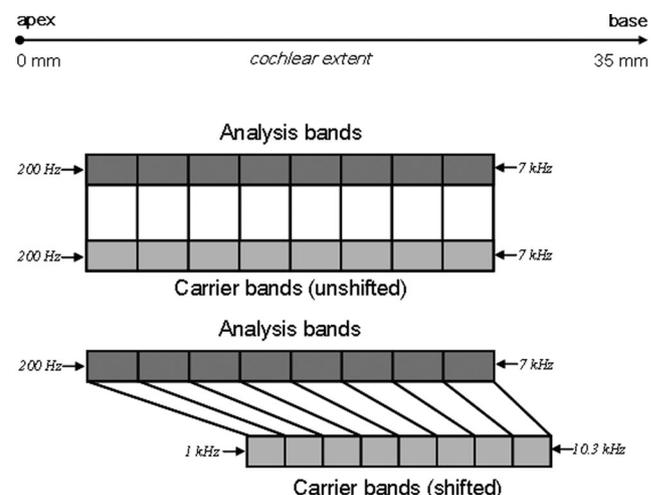
implemented as follows. The signal was first processed through a pre-emphasis filter (high-pass with a cutoff frequency of 1200 Hz and a slope of 6 dB/octave). The input frequency range (200 to 7000 Hz) was bandpass-filtered into eight spectral bands using 4th-order Butterworth filters. The corner frequencies of the bandpass filters were calculated according to Greenwood's (1990) formula; thus, each bandpass filter was comparable in terms of cochlear extent. The corner frequencies (3 dB down) of the analysis filters are listed in Table 1. The temporal envelope was extracted from each frequency band by half-wave rectification and low-pass filtering at 160 Hz. The extracted envelopes were used to modulate sinusoidal carriers. The modulated carriers of each band were summed and the overall level was adjusted to be the same as the original speech. The frequencies of the carriers depended on the experimental condition. For the spectrally unshifted condition, the frequencies of the sine wave carriers were equal to the center frequencies of the analysis filters. For the spectrally shifted condition, the carrier frequency bands were upwardly shifted to simulate a shallow insertion of a 16-mm-long, eight-electrode array with 2-mm electrode spacing; sine wave carrier frequencies were equal to the center frequencies of the shifted carrier bands. The analysis and carrier filters, sine wave carrier frequencies, and cochlear distance from the apex (according to Greenwood, 1990) are shown in Table 1, for both unshifted and shifted speech. Note that for the shifted speech condition, the output signal was both spectrally shifted and compressed to simulate two aspects of spectral distortion typically associated with cochlear implant devices and speech processing. Figure 1 illustrates the distribution of analysis and carrier bands for both the unshifted and shifted speech conditions.

### C. Test and Training Materials

Speech recognition was assessed with the use of multitalker vowel, consonant, and sentence recogni-

tion. The vowel test stimuli included 12 medial vowel tokens presented in a /h/-vowel-/d/ context ("heed," "hid," "hayed," "head," "had," "hod," "hawed," "hoed," "hood," "who'd," "hud," and "heard"). Vowel tokens were digitized natural productions from five male and five female talkers, drawn from speech samples collected by Hillenbrand et al. (1995). Consonant stimuli included 20 medial consonant tokens presented in an /a/-consonant-/a/ context ("aba," "ada," "aga," "apa," "ata," "aka," "ala," "ara," "aya," "awa," "ama," "ana," "afa," "asa," "asha," "ava," "aza," "atha," "acha," and "aja"). Consonant tokens were digitized natural productions from five male and five female talkers (recorded by Shannon et al., 1999). Sentences were digitized natural productions from the IEEE database (1969) from one male and one female talker, recorded at House Ear Institute; 72 lists of 10 sentences each were available for testing.

Training stimuli included more than 1,000 monosyllabic words and were digitized natural productions from two male and two female talkers (Fu et al., 2005a). The talkers used for the training stimuli



**Fig. 1.** Frequency allocations of analysis and carrier filter bands for 8-channel acoustic simulations of cochlear implant speech processing.

were not the same as those used for the test stimuli, with the exception of one male talker, who was used for both the training stimuli and the IEEE sentence stimuli.

#### D. Test and Training Procedures

Test and training materials were presented at 65 dBA in free field in a double-walled, soundproof booth through a Tannoy speaker. Vowel and consonant recognition was measured before, during, and after training. Sentence recognition was measured before and after training. Vowel recognition was measured in a 12-alternative identification paradigm. Consonant recognition was measured in a 20-alternative identification paradigm. Sentence recognition was measured in an open-set recognition paradigm. For phoneme testing, vowel stimuli included 120 tokens (12 vowels \* 10 talkers); consonant stimuli included 200 tokens (20 consonants \* 10 talkers). During each trial of a phoneme recognition test, a stimulus token was chosen randomly, without replacement, and presented to the subject. The subject responded by clicking on one of the response buttons shown on-screen (12 response buttons for the vowel test, 20 response buttons for the consonant test). The response buttons were labeled in a /h/-vowel-/d/ context for vowel recognition and /a/-consonant-/a/ context for consonant recognition. No feedback was provided, and subjects were instructed to guess if they were not sure, although they were cautioned not to provide the same response for each guess. For sentence testing, two lists were chosen from among the 72 sentence lists. Sentence recognition was measured for each list, spoken by a different talker, for a total of 20 sentences; thus, sentence List 1 would be tested for male Talker 1 and sentence List 2 would be tested with female Talker 2. During each trial of the sentence recognition test, a sentence would be chosen randomly, without replacement, from among the 10 sentences in the test list. The subject responded by repeating as many words as possible, and the experimenter scored the number of correctly identified words in each sentence. No feedback was provided.

Baseline phoneme and sentence recognition was measured before training was begun. Phoneme and sentence recognition was first tested by using unprocessed speech to familiarize subjects with the test tokens, labels, and formats and to ensure that subjects were capable of near-perfect recognition of the unprocessed speech stimuli. After baseline testing with unprocessed speech, vowel, consonant and sentence recognition was tested for 8-channel, spectrally unshifted sine wave speech. Baseline measures with spectrally unshifted speech were re-

peated three times for vowel recognition (or until performance asymptoted), and two times for consonant and sentence recognition. After testing with the 8-channel unshifted speech, vowel, consonant and sentence recognition was tested for 8-channel, spectrally shifted sine wave speech. Baseline measures with spectrally shifted speech were repeated three times for vowel recognition (or until performance asymptoted), two times for consonant recognition and three times for sentence recognition. After baseline measures were obtained, subjects completed five training sessions. Immediately before and after each training session, vowel and consonant recognition with shifted speech was retested. Each training session lasted 1 hr. At the end of the fifth and final training session, vowel, consonant, and sentence recognition with shifted speech was retested. As with the baseline measures, pretraining measures with shifted speech were repeated three times for vowel recognition, two times for consonant recognition, and three times for sentence recognition. Vowel, consonant, and sentence recognition was also retested with unshifted speech to verify that any improvement in performance due to training was due to learning the spectral shift and compression, rather than simply learning 8-channel sine wave processing.

The 18 subjects were divided into three groups of six subjects each. Each group was trained at one of three rates: one session per week (1× group), three sessions per week (3× group), or five sessions per week (5× group). Note that in the 5× group, data for 4 of the 6 subjects were previously reported in Fu et al. (2005b). There were no sentence recognition data for those 4 subjects. There were 5 total training sessions for each group. No control group was used due to previous results in Fu et al. (2005b), which showed no significant improvement in vowel recognition for subjects who received no training but repeated vowel tests daily for 5 consecutive days.

Targeted vowel contrast training was conducted by using the protocol described in Fu et al. (2005a, 2005b). Subjects were trained with the use of custom software (Computer-Assisted Speech Training, or CAST, developed at House Ear Institute) and monosyllabic cVc words; the training stimuli were produced by a different set of talkers than used for test stimuli. Training stimuli were processed exactly the same as the shifted speech test stimuli. During training, a stimulus was presented to the subject. Depending on the level of difficulty, there were two, four, or six response choices; only the medial vowel differed between response choices (i.e., “seed,” “said”), allowing subjects to focus on medial vowel differences. Initially, the response choices differed greatly in terms of acoustic speech features (i.e., “said,”

**TABLE 2.** Mean vowel, consonant, and sentence recognition scores before and after training with shifted speech

	Vowel (% correct)		Consonant (% correct)		Sentence (% correct)	
	Pretrain	Posttrain	Pretrain	Posttrain	Pretrain	Posttrain
Unprocessed speech						
All subjects	93.4		97.7		100.0	
1×/wk only	95.2		98.1		100.0	
3×/wk only	93.1		97.1		100.0	
5×/wk only	91.9		97.9		99.7	
8-channel unshifted speech						
All subjects	82.4	84.0	90.9	91.0	96.8	96.4
1×/wk only	87.5	87.4	90.7	90.5	97.4	96.3
3×/wk only	79.7	82.9	91.0	92.1	95.5	95.9
5×/wk only	81.9	84.2	90.9	90.4	98.7	97.9
8-channel shifted speech						
All subjects	14.0	29.3	44.4	64.6	30.2	54.7
1×/wk only	14.9	28.3	46.5	63.1	32.5	55.8
3×/wk only	13.7	28.5	41.9	63.7	26.3	51.3
5×/wk only	13.3	31.3	44.9	67.0	35.4	61.5

Note that measures with unprocessed speech were taken during baseline testing only and were not repeated after training. Mean vowel, consonant, and sentence recognition scores before and after training with shifted speech.

“sued”); as subjects’ performance improved, the difference in speech features among the response choices was reduced (i.e., “said,” “sad”). The acoustic speech features used to define these levels of difficulty included tongue height (which is associated with the acoustic frequency of the vowel first formant, or, F1), tongue position (from front to back of the oral cavity, which is associated with the frequency difference between F2 and F1) and vowel duration. As subjects’ performance improved beyond a criterion level (80% correct), the number of response choices was increased and/or the acoustic feature differences between the response choices was reduced. For example, subjects started with two choices, and vowel feature contrasts were decreased from high to low in three steps; the next three levels had four choices, with vowel feature contrasts decreasing from high to low. Audio and visual feedback was provided. If the subject responded correctly, visual feedback was provided and a new stimulus was selected. If the subject responded incorrectly, auditory and visual feedback was provided; the correct response and the subjects’ (incorrect) response were played in sequence repeatedly, allowing subjects to directly compare the two choices. Each training block contained 50 trials. Subjects completed as many training blocks as they could within each 1-hr session.

## RESULTS

Note that for all figures, and within the text, data are reported as percent correct or as the shift in performance in percentage points. For statistical analyses, subject scores were transformed to rationalized arcsine units (rau) (Studebaker, 1985) to

correct for floor and ceiling effects in subjects’ performance. Table 2 summarizes the results for all measures, listed in percent correct.

Near-perfect recognition scores were obtained for all subjects in the preliminary vowel, consonant, and sentence recognition tests with unprocessed speech. When the spectral resolution was reduced (8-channel unshifted speech), mean vowel recognition dropped 11 percentage points, mean consonant recognition dropped 7 percentage points, and mean sentence recognition only dropped about 3 percentage points. When the signal was both spectrally degraded and shifted (8-channel, spectrally shifted speech), mean vowel recognition dropped almost 80 percentage points from the unprocessed speech scores to about chance level, mean consonant recognition dropped 44 percentage points, and mean sentence recognition dropped 45 percentage points.

Figure 2 shows mean vowel, consonant, and sentence recognition scores with 8-channel spectrally shifted speech for the three training groups, before and after training. For all groups, and for all speech measures, mean performance improved with the vowel contrast training. A two-way ANOVA, with training and training rate as factors, showed a significant main effect of training on vowel recognition [ $F(1,30) = 25.585, p < 0.001$ ], consonant recognition [ $F(1,30) = 30.004, p < .001$ ], and sentence recognition [ $F(1,22) = 21.695, p < 0.001$ ]. However, there was no significant effect of training rate on either vowel recognition [ $F(2,30) = 0.039, p = 0.962$ ], consonant recognition [ $F(2,30) = 0.279, p = 0.759$ ], or sentence recognition [ $F(2,22) = 1.194, p = 0.322$ ]. Although the trend in Figure 2 shows greater training effects for the 3× and 5× groups than for the 1× group, statistical analysis revealed that for training rate effects, power (with  $\alpha = 0.05$ ) was only 0.05 for vowels and consonants and 0.0746 for sen-

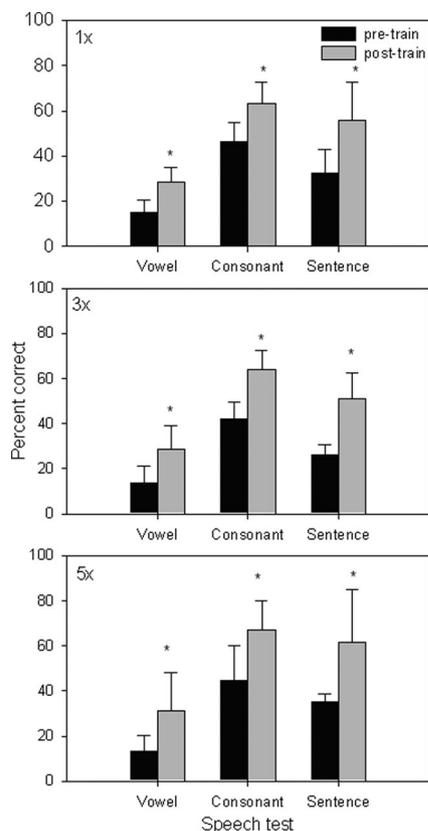


Fig. 2. Mean pre- and posttraining vowel, consonant recognition and sentence recognition scores for 8-channel, spectrally shifted speech. From top to bottom, results are shown for three training rates: one session per week (1x), three sessions per week (3x), and five sessions per week (5x). Asterisks indicate a significant difference between pre-and posttraining performance. Error bars indicate  $\pm 1$  standard deviation. It should be noted that sentence recognition data from the 5x group were limited to only 2 of the 6 subjects.

tences. The power analysis further revealed that the required difference between groups to achieve sufficient statistical power (power >0.8) was approximately 19% for vowels, 21% for consonants, and 24% for sentences.

Figure 3 shows mean and individual shift in vowel recognition performance (in percentage points), relative to pretraining performance, as a function of training session. Note that for the 5x group, the data shown for subjects S13-S16 were previously reported in Fu et al. (2005b). Although there seems to be a trend that shows more accelerated improvement with increased training rate, there were no significant differences between the three training rates. A two-way ANOVA (performed on the raw data, including baseline performance), with training rate and training session as factors, showed a significant main effect of training session [ $F(5,90) = 5.509, p < 0.001$ ] but no significant effect of training rate [ $F(2,90) = 0.0911, p = 0.913$ ]. Post hoc Tukey pairwise comparisons showed

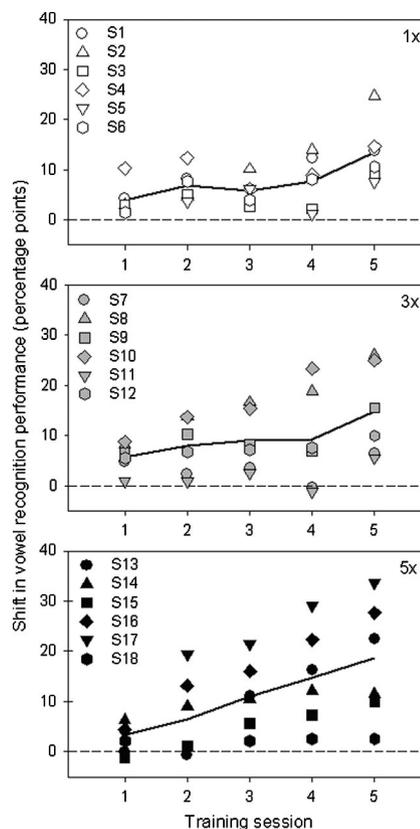


Fig. 3. Shift in vowel recognition performance (in percentage points), relative to pretraining performance, as a function of training session. From top to bottom, results are shown for three training rates: one session per week (1x), three sessions per week (3x), and five sessions per week (5x). Individual data are shown by different symbols; mean data are shown by solid lines.

that there was no significant improvement in performance until the third training session.

Performance with 8-channel, unshifted speech was retested after training was completed. Since there was no significant difference across the training groups, all data from the different training groups were grouped together (Table 2). One-way ANOVA tests showed no significant effect of training on unshifted vowel recognition [ $F(1,34) = 0.480, p = 0.493$ ], consonant recognition [ $F(1,34) = 0.055, p = 0.817$ ], or sentence recognition [ $F(1,26) = 0.163, p = 0.690$ ].

### DISCUSSION

The results of the present study demonstrate that moderate amounts of auditory training can significantly improve recognition of spectrally shifted speech, consistent with previous studies (Fu & Galvin, 2003; Fu et al., 2005b; Rosen et al., 1999). The results showed no significant difference in training outcomes when subjects trained one to five times per week, when a total of five training sessions was

completed. It should be noted that the designation of five training sessions was arbitrary. It is not implied that performance will plateau after five training sessions, nor that difference between training rates may become apparent over a longer training period. The present study also offers several interesting findings regarding effects of auditory training on the recognition of spectrally shifted speech.

Although auditory training significantly improved mean vowel recognition scores, there was a large intersubject variability in training outcomes for all three training groups. Individual subjects improved from a minimum of 3 to a maximum of 34 percentage points after completing 5 training sessions. There was also large intersubject variability in terms of the time course of improvement. Some subjects improved incrementally after each training session, whereas others showed no improvement during the first few training sessions. The large intersubject variability observed in the present study may be explained in terms of NH subjects' motivation to learn. Because of the extreme spectral shift and compression, most subjects could understand little if any of the speech presented before training. For most people, this was very discouraging. NH subjects were only exposed to the cochlear implant simulations during the training and test sessions, after which they would return to their normally hearing lives. Whether or not they improved their performance with the shifted speech, they were still paid at the end of the experiment. On the other hand, cochlear implant listeners will never return to a normal hearing world. If they can improve their understanding of the spectrally distorted speech, then they will be able to communicate with others and perceive what is going on in their environment. Since communication in the real world is at stake, subjects with cochlear implants may be more uniformly motivated to learn spectrally shifted speech. It is also possible that the vowel contrast training protocol used in the present study may only be suitable for some subjects. Different training protocols may be required for different subjects to effectively adapt to spectrally shifted speech.

Surprisingly, the frequency of training had no significant effect on training outcomes, at least within the experimental training period. For the 1× group, the improvement in vowel recognition ranged from 7.6 to 24.7 percentage points, with a mean of 13.4 percentage points. For the 3× group, the improvement in vowel recognition ranged from 5.6 to 26.1 percentage points, with a mean of 14.8 percentage points. For the 5× group, the improvement ranged from 2.6 to 33.6 percentage points, with a mean of 18.0 percentage points. Because of the large intersubject variability, there was no significant

difference among the training groups in terms of the time course or amount of improvement. However, careful examination of the data shows some potential trends in terms of training rate. First, the mean improvement for the 5× group was ~5 percentage points higher relative to the 1× group, and ~3 points higher relative to the 3× group. Second, for the 5× group, 3 of the 6 subjects improved more than 20 percentage points after completing 5 training sessions; only 1 subject in the 1× group and 2 subjects in the 3× group improved by 20 percentage points or more. Again, subject motivation may have contributed to these trends in results. If all NH subjects experienced the urgency of learning that probably is experienced by patients with cochlear implants, it is possible that the frequency of training may have significantly affected the training outcomes in the present study.

Another interesting finding is that the improved vowel recognition with the vowel contrast training protocol generalized to improved consonant and sentence recognition, for all three training groups. Note that different stimuli and talkers were used for training and testing, and that consonant and sentence recognition was not explicitly trained; training was performed using more than 1000 monosyllabic words in a cVc format. All 20 of the consonants in the consonant recognition tests were present in the cVc training words as initial and final consonants, though only those combinations of consonants and vowels that could create commonly used words were used in the training word database. The improvement in consonant and sentence recognition was comparable among the three training groups. The improved recognition of spectrally shifted consonants suggests that subjects benefited from exposure to the initial and final consonants in the monosyllabic training words, consistent with results from a previous training study (Fu et al., 2005b). In that Fu et al. study, all subjects, who were exposed to the shifted consonants during consonant tests, improved their shifted consonant recognition; subjects who had additional exposure to consonants, through targeted vowel training with the monosyllabic cVc words or through sentence training, improved their shifted consonant recognition even more. Although consonant discrimination or recognition was not directly trained in the targeted vowel training with monosyllabic words, the exposure to the consonant sound in conjunction with the visual word labels helped improve shifted consonant recognition.

After training with spectrally shifted speech was completed, retesting with 8-channel spectrally unshifted speech showed no significant difference in vowel, consonant, and sentence recognition from baseline measures. Thus, training with spectrally

shifted speech did not seem to generalize to improved performance for frequency carrier ranges other than those used for training, consistent with results from previous studies (Fu & Galvin, 2003; Fu et al., 2005b). However, baseline performance with 8-channel unshifted speech was already at a high level, leaving little room for improvement. Given that subjects were trained to listen to speech that was both spectrally reduced and shifted (relative to unprocessed speech), it seems unlikely that training would have improved recognition of spectrally reduced speech, as this parameter had the smallest effect on baseline performance. Rosen et al. (1999) found a slight improvement in unshifted 4-channel speech after training with shifted 4-channel speech, but the change was small compared with the improvement with shifted 4-channel speech. This small increase in performance for unshifted 4-channel speech may have been related to “procedural” learning as opposed to training. To avoid as much of this effect as possible in our study, baseline tests for each condition were repeated until performance asymptoted, as opposed to a fixed number of runs of initial baseline tests (two, in the case of Rosen et al. (1999) for spectrally reduced 4-channel speech). Also, it is unlikely that subjects in the present experiment experienced other “procedural” types of learning. In Fu et al. (2005b), performance from a control “test only” group was compared with that of the “vowel contrast training” group; note that data from the 4 subjects in vowel training group were included in the present study (5× group). There was no significant difference in performance for the test-only group after 5 consecutive days of testing, suggesting that the improved vowel recognition in the vowel training group was due to “perceptual” learning rather than procedural learning (Hawkey et al., 2004; Wright & Fitzgerald, 2001).

These results, combined with those from previous studies (Fu and Galvin, 2003; Fu et al., 2005a, 2005b; Rosen et al., 1999), suggest that auditory training may be an effective approach toward improving patients with cochlear implants’ speech recognition. The present study also suggests that the amount of training rather than the frequency of training may strongly influence patients with cochlear implants’ training outcomes. Given that the present study was conducted with NH subjects listening to cochlear implant simulations, it should also be noted that although it is possible to simulate cochlear implant speech processing, it may not always be possible to simulate the urgency of the learning process experienced by patients with cochlear implants. As such, the effect of training rate may be somewhat different with patients with cochlear implants.

## SUMMARY AND CONCLUSION

The present study showed that moderate amounts of auditory training significantly improved NH listeners’ recognition of spectrally shifted speech. Completing five training sessions at three different training frequencies revealed:

1. Targeted vowel contrast training using monosyllabic words significantly improved recognition of spectrally shifted vowels, even with only one training session per week.
2. There was no significant difference in training outcomes when subjects trained one, three, or five times per week, when five training sessions in total were completed.
3. For spectrally shifted speech, the improved vowel recognition performance with the vowel contrast training protocol generalized to improved recognition of consonants and sentences.
4. There was large intersubject variability in terms of the amount and time course of improvement with training, suggesting that individualized training protocols may be appropriate for different subjects.

## ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in our experiments. Research was supported by NIDCD grant R01-DC004792.

Address for correspondence: Dr. Qian-Jie Fu, Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057. E-mail: qfu@hei.org.

Received February 21, 2006; accepted September 2, 2006.

## REFERENCES

- Busby, P. A., Roberts, S. A., Tong, Y. C., Clark, G. M. (1991). Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. *British Journal of Audiology*, 25, 291–302.
- Dawson, P. W., Clark, G. M. (1997). Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear and Hearing*, 18, 488–501.
- DeFilippo, C. L., Scott, B. L. (1978). A method for training and evaluation of the reception of on-going speech. *Journal of the Acoustical Society of America*, 63, 1186–1192.
- Fitzgerald, M. B., Wright, B. A. (2005). A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *Journal of the Acoustical Society of America*, 118, 3794–3803.
- Fu, Q.-J., Galvin, J. J. (2003). The effects of short-term training for spectrally mismatched noise-band speech. *Journal of the Acoustical Society of America*, 113, 1065–1072.
- Fu, Q.-J., Galvin, III, J. J., Wang, X., Nogaki, G. (2005a). Moderate auditory training can improve speech performance of adult cochlear implant users. *Acoustics Research Letters Online*, 6, 106–111.
- Fu, Q.-J., Nogaki, G., Galvin, J. J. (2005b). Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology*, 6, 180–189.

- Greenwood, D. D. (1990). A cochlear frequency-position function for several species: 29 years later. *Journal of the Acoustical Society of America*, *87*, 2592–2605.
- Hawkey, D. J., Amitay, S., Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience*, *7*, 1055–1056.
- Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–3111.
- IEEE (1969). IEEE recommended practice for speech quality measurements. Institute of Electrical and Electronic Engineers, New York, 1969.
- Rosen, S., Faulkner, A., Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *Journal of the Acoustical Society of America*, *106*, 3629–3636.
- Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., Wang, X. (1999). Consonant recordings for speech testing. *Journal of the Acoustical Society of America*, *106*, L71–L74.
- Studebaker, G. A. (1985). A 'rationalized' arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455–462.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., Rabinowitz, W. M. (1991). New levels of speech recognition with cochlear implants. *Nature*, *352*, 236–238.
- Wright, B. A., Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Science*, *98*, 12307–12312.

#### REFERENCE NOTE

- Ortiz J. A., Wright B. A., Fitzgerald M. B., Pillai, J. (2001). Rapid improvements on interaural-time-difference discrimination: Evidence for three types of learning. *Journal of the Acoustical Society of America*, *109*, 2289(A).